

## Title

ARITHMETIC UNIT

## Field of the Invention

10 The present invention relates to an arithmetic unit and in particular but not exclusively to an arithmetic unit for use in a digital signal processor.

## Background to the Invention

15 In known arithmetic units, it is possible to calculate a multiply accumulate function. In a multiply accumulate function, a first number X is multiplied by a second number Y and added to an accumulator ACCU. This function may be repeated a number of times  
 20 to modify each time the value of the accumulator ACCU. In mathematical notation, the result RES of this function can be represented as follows:-

$$\text{RES} = \text{ACCU} + X * Y$$

25 In the next operation which is performed, ACCU will equal the result RES obtained by carrying out the preceding function.

Reference is made to Figure 1 which shows a known arithmetic unit  
 30 2 which performs a multiply accumulate function. The unit 2 generates n partial products 4. The current accumulator ACCU is stored in the accumulator register 6. The partial products 4 along with the output from the accumulator register 6 are input to a carry save adder 8. The carry save adder 8 reduces the  
 35 number of partial products and the output of the accumulator register 6 to two partial products. In other words, the n+1 inputs to the carry save adder 8 are reduced to two. These two partial products are output to a final adder 10 which adds together the two partial products to provide the final result  
 40 RES.

5 In the arrangement shown in Figure 1, it is possible to perform a carry save operation in the carry save adder 8 on the partial products and the contents of the accumulator register 6 at the same time.  $X*Y$  is equal to the sum of all of the partial products  $P_0$  to  $P_{(n-1)}$ . In other words  $X*Y = \sum P_i$  for  $i=0,1\dots n-1$ . Thus,

10  $RES=ACCU+\sum P_i$ . As shown by this equation, all the partial products and the accumulator  $ACCU$  are added together in order to obtain the result  $RES$ .

15 It is also desirable in certain circumstances to be able to perform a multiply subtract operation. In this operation, a first number  $X$  is multiplied by a second number  $Y$  and subtracted from a accumulator  $ACCU$  (which in fact may be modified by subsequent operations) to provide the result  $RES$ . In mathematical notation,  $RES=ACCU-X*Y$ . Again,  $X*Y$  can be determined by adding together the

20  $n$  partial products. Accordingly,  $RES=ACCU-\sum P_i$ . The architecture shown in Figure 1 is not suitable for use with a multiply subtract operation. This is because the partial products and accumulator  $ACCU$  cannot simply be added together to provide the result  $RES$ . Some additional circuitry will therefore be required

25 in order to allow a multiply subtract operation to be performed as well as a multiply accumulate function. This additional circuitry would inevitably result in an increase in the time taken to complete an operation.

### 30 Summary of the Invention

It is therefore an aim of embodiments of the present invention to provide an arithmetic unit which avoids or at least mitigates the difficulties of the above discussed arrangement.

35

According to one aspect of the present invention there is provided an arithmetic unit for multiplying a first quantity  $X$  by a second quantity  $Y$ , said arithmetic unit comprising:

40 a Booth coder having a plurality of inputs for receiving a plurality bits of the second quantity and a plurality of outputs for providing Booth coded outputs; and

10

10

15

25

35

40

5 products to be reduced by 2.

Three multiplexers may be provided, the first and second multiplexers each receiving the NZP and NZN outputs of the first and second Booth coders and the third multiplexer receiving the  
 10 SNGL outputs of the first and second Booth coders. In this type of arrangement, multiplexers are in any event required in order to select the most significant bits or the least significant bits. Accordingly, the use of the multiplexers to provide outputs in accordance with whether or not the partial product or the  
 15 partial product of the opposites sign is required does not necessarily increase the time taken to complete an operation.

#### Brief Description of the Drawings

20 For a better understanding of the present invention and as to how the same may be carried into effect, reference will now be made by way of example only to the accompanying drawings in which:

Figure 1 shows a block diagram of a known arithmetic unit which  
 25 is arranged to perform a multiply accumulate function;

Figure 2 shows a block diagram illustrating how Booth coding works;

Figure 3 shows a block diagram of a modified Booth coder embodying the present invention;

30 Figure 4 shows a block diagram of an arrangement incorporating conventional Booth coding, with a selection between the most significant bits and the least significant bits;

Figure 5 shows a block diagram incorporating Booth coding which embodies the present invention, with a selection between the most  
 35 significant bits and the least significant bits;

Figure 6 shows a second Booth coder embodying the present invention;

Figures 7 and 8 show a further embodiment of the present invention.

## 5 Detailed Description of Embodiments of the Invention

Embodiments of the present invention are arranged to be able to carry out multiply accumulate and multiply subtract functions. In the multiply subtract function,  $-P_i$  is used instead of  $P_i$ .  $P_i$  can be positive or negative. If  $P_i$  is positive, then  $-P_i$  will be negative and vice versa. In other words, the opposite value of each partial product is generated. These opposite partial products  $-P_i$  can then be added together with the accumulated total ACCU as in the arrangement shown in Figure 1 for the multiply accumulate function.

In a preferred embodiment of the present invention,  $-P_i$  is generated by using a modified form of Booth coding. Booth coding is a well known algorithm which is sometimes referred to as the Booth-MacSorley algorithm. It will be referred to simply as the Booth algorithm in this document. The Booth algorithm is used to recode the multiplier (i.e.  $Y$ ) such that the number of partial products is roughly reduced by a factor of 2. The multiplier, i.e.  $Y$  is divided into groups of three bits, each group comprising the  $2i+1$ , the  $2i$  and  $2i-1$  bits of the  $Y$  multiplier. Depending on the values of the bits of each group, the value of the number to be multiplied by  $Y$ , i.e.  $X$  is modified to generate one of five partial products:  $-2X$ ,  $-X$ ,  $0$ ,  $X$ ,  $2X$ .

30 Reference is made to the following table, table 1, which shows the partial products for all the possible values of the group of three bits of the multiplier Y.

5

Y2i+1	Y2i	Y2i-1	Partial Product	Coding SNGL-NZP-NZN The carry is NZN			Coding CX-C2X-SGN The carry is SGN		
				SNG L	NZP	NZN	CX	C2X	SGN
0	0	0	0	0	0	0	0	0	0
0	0	1	X	1	1	0	1	0	0
0	1	0	X	1	1	0	1	0	0
0	1	1	2X	0	1	0	0	1	0
1	0	0	-2X	0	0	1	0	1	1
1	0	1	-X	1	0	1	1	0	1
1	1	0	-X	1	0	1	1	0	1
1	1	1	0	0	0	0	0	0	1

This Booth coding can be used in two different ways. In one form of coding referred to as SNGL-NZP-NZN coding, the carry is NZN. A first output SNGL will have the value "1" when the partial product is +X or -X. The remaining values will be "0". NZP is the "non zero positive" value and has the value "1" when the partial product is X or 2X. At other times, the value will be "0". NZN is the "non zero negative" value and has the value "1" if the partial product is -X or -2X. At other times the value will be "0".

In the other form of coding referred to as CX-C2X-SGN, the carry signal is SGN. CX will have the value "1" when the partial product is X or -X and at other times the value will be "0". C2X will have the value "1" when the partial product is equal to +2X



5 or  $-2X$  and at other times the value will be "0". SGN will have the value "1" if the partial product is  $-X$  or  $-2X$  and in one of the 2 cases where the partial product is 0 (when the 3 input bits are all equal to "1"). In other cases, its value is "0".

10 In both of the Booth coding examples discussed hereinbefore, 2's complement arithmetic is used.

Reference will now be made to Figure 2 which illustrates how Booth coding works in practice.

15 The three bits of the multiplier  $Y_{2i+1}$ ,  $Y_{2i}$  and  $Y_{2i-1}$  are input to a Booth coder. In practice a number of Booth coders are provided. In dependence of the value of the individual bits of the multiplier  $Y$  input into the Booth coder 20, an output of three bits is provided. The output provided by the Booth coder 20 will depend also on the form of coding as discussed hereinbefore.

25 The three bits which are output by the Booth coder are input to a Booth decoder 22. The Booth decoder 22 comprises a plurality of Booth decoder units 24. The number of Booth decoder units 24 is roughly equal to the number of bits in the multiplicand. Each Booth decoder unit 24 receives an input from two bits of the multiplicand,  $X_j$  and  $X_{j-1}$ . Each Booth decoder unit 24 provides a partial product output. The carry bit is provided by the NZN or 30 SGN bits depending on the type of Booth coding used.

In embodiments of the present invention,  $-P$  is obtained by carrying out the following operation:  $-P = X * (-Y)$ . In 2's complement arithmetic,  $-Y = \bar{Y} + 1$  where  $\bar{Y}$  is the complement of  $Y$ . Accordingly, in the Booth coder, instead of  $Y_{2i+1}$ ,  $Y_{2i}$  and  $Y_{2i-1}$ , 35 the following need to be considered:

$$\overline{Y_{2i+1}}, \overline{Y_{2i}}, \overline{Y_{2i-1}}$$

The carry i.e. +1 can be dealt with separately.

100E70-3646T60

- 5 Reference is made to the modified Booth encoding provided by embodiments of the present invention. First of all, the modification required to the SNGL-NZP-NZN coding will first be discussed. Reference is made to table 2.

Y2i+1	Y2i	Y2i-1	Negative Product MUL_SUB	Partial Product	SNGL'	NZP'	NZN'
0	0	0	0	0	0	0	0
0	0	1	0	X	1	1	0
0	1	0	0	X	1	1	0
0	1	1	0	2X	0	1	0
1	0	0	0	-2X	0	0	1
1	0	1	0	-X	1	0	1
1	1	0	0	-X	1	0	1
1	1	1	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	1	1	-X	1	0	1
0	1	0	1	-X	1	0	1
0	1	1	1	-2X	0	0	1
1	0	0	1	2X	0	1	0
1	0	1	1	X	1	1	0
1	1	0	1	X	1	1	0
1	1	1	1	0	0	0	0

The first three columns represent the possible values of three bits of the multiplier i.e. Y. The fourth column represents



5 whether or not a multiply accumulate or a multiply subtract operation is to be carried out. If a multiply accumulate function is to be carried out, then  $P$  is calculated by the multiplier and  $MUL\_SUB$  is equal to "0". For a multiply subtract function,  $-P$  is calculated and  $MUL\_SUB$  is equal to "1". The value of  $MUL\_SUB$  thus indicates if a multiply accumulate or a multiply subtract operation is to be performed. The first half of the table is thus the same as table 1 and represents the Booth coding values for the multiply accumulate function whilst the second part of the table represents the values for the multiply subtract function.

10 In a multiply subtract function, the partial product will have the opposite sign to the partial product for a multiply accumulate function for the same values of the group of three bits of the multiplier  $Y$ .

20 From a comparison of the first half of the table it can be seen that for given values of the three bits of the multiplier  $Y$ , the output  $SNGL'$  is unaffected by whether or not a multiply accumulate or subtract function is to be performed. It can also be seen that for given values of the three bits of the multiplier  $Y$ ,  $NZP'$  for a multiply accumulate function equals  $NZN'$  for a multiply subtract function. Likewise, for given values of the three bits of the multiplier  $Y$ ,  $NZN'$  for a multiply accumulate function equals  $NZP'$  of the multiply subtract function.

25

30 Reference is made to Figure 3 which shows a Booth coder 32 embodying the present invention. The Booth coder 32 embodying the present invention comprises a conventional Booth coder 20 as shown in Figure 2. However, the  $NZP$  output is provided to first and second multiplexers 26 and 28 respectively. Likewise, the  $NZN$  output is also provided to the first and second multiplexers 26 and 28. The first and second multiplexers 26 and 28 each receive a control signal 30 which has the value "0" if a multiply accumulate function is to be performed and the value "1" if a multiply subtract function is to be performed. The output of the first multiplexer 26 provides the  $NZP'$  output whilst the output of the second multiplexer 28 provides the  $NZN'$  output.

35

40



5 In accordance with the normal Booth coding principles, the signals NZP, NZN and SNGL are generated by the first and second Booth coders 40 and 42. The outputs of the first and second Booth coders 40 and 42 are output to first, second and third multiplexers 44, 46 and 48 respectively. The first multiplexer 44 receives the NZP signal from both the first and the second Booth coders 40 and 42. The second multiplexer 46 receives the NZN signal from both the first and the second Booth coders 40 and 42 and the third multiplexer 48 receives the SNGL signal from the first and second Booth coders 40 and 42. The multiplexers 44, 46 and 48 receive a control signal 50. If the control signal 50 has the value "1", then the most significant bits are to be used in the multiplication. If, on the other hand, the control signal 50 has the value zero, then the least significant bits are to be used for the multiplication.

20 Thus, if the control signal 50 has the value "1", the outputs from the first Booth coder 40 are output from the first to third multiplexers 44 to 48 respectively. If, on the other hand, the least significant bits are to be used for the multiplication and the control signal 50 has the value "0", then the NZP, NZN and SNGL outputs from the second Booth coder 42 will be output by the first to third multiplexers 44 to 48 respectively.

30 Reference is made to Figure 5 which shows how embodiments of the present invention can be used in an arrangement such as shown in Figure 4. The Booth coder unit 60 embodying the present invention comprises first and second Booth coders 62 and 64. The first and second Booth coders 62 and 64 are conventional. The first Booth coder 62 receives the most significant bits and therefore the inputs are  $Y_{2i+N+1}$ ,  $Y_{2i+N}$  and  $Y_{2i+N-1}$ . The second Booth coder 64 receives the least significant bits and therefore receives the inputs  $Y_{2i+1}$ ,  $Y_{2i}$  and  $Y_{2i-1}$ . The first and second Booth coders 62 and 64 each generate the conventional NZP, NZN and SNGL signals as discussed hereinbefore.

40 The Booth coder unit 60 embodying the present invention has first

- 5 to third multiplexers 66, 68 and 70. The first multiplexer 66 receives inputs from the NZP and NZN outputs of both the first and second Booth coders 62 and 64. In other words, the first multiplexer 66 receives four inputs.
- 10 The second multiplexer 68 receives the same inputs as the first multiplexer 66 and therefore receives the NZP and NZN outputs from both the first and second Booth coders 62 and 64. The first and second multiplexers 66 and 68 each receive four control signals 72 to 78. The first control signal 72 indicates whether a
- 15 subtract function is to take place using the least significant bits. The second control signal 74 indicates whether a accumulate function is to be carried out with the least significant bits. The third control signal 76 indicates whether a subtract function is to be carried out with the most significant bits whilst the
- 20 fourth control signal 78 indicates if an accumulate function is to be carried out with the most significant bits. Accordingly, one of these signals will have the value "1" whilst the rest of the signals will have the value "0".
- 25 If the first control signal 72 has the value "1", then the NZN output of the second Booth coder 64 will provide the output of the first multiplexer 66 which will be the NZP' output. If the value of the second control signal 74 is "1", then the output provided by the first multiplexer 66 will be the NZP signal from
- 30 the second Booth coder 64. If the third control signal 76 has the value "1", the NZN output of the first Booth coder 62 will be output by the first multiplexer 66 to provide the NZP' output. Finally, if the fourth control signal 78 has the value "1", the output of the first multiplexer 66 will be the NZP output of the
- 35 first Booth coder 62.

These same control signals also control the output of the second multiplexer 68 which provides the NZN' output. If the value of the first control signal 72 is "1", then the output of the second

40 multiplexer 68 will be the NZP output of the second Booth coder 64. If the value of the second control signal 74 is "1" then the

09019196-073001  
TOP SECRET

5 output of the second multiplexer 68 will be the NZN output of the  
 second Booth coder 64. If the value of the third control signal  
 76 is "1", then the output of the second multiplexer 68 will be  
 the NZP output of the first Booth coder 62. Finally, if the value  
 of the fourth control signal 78 is "1", then the output of the  
 10 second multiplexer 68 will be the NZN output of the first Booth  
 coder 62.

The third multiplexer 70 receives the SNGL signals from the first  
 and second Booth coders 62 and 64. The third multiplexer 70  
 15 receives a control signal 71 which selects the output from the  
 first Booth coder 62 if the most significant bits are to be used  
 in the multiplication and the output of the second Booth coder 64  
 if the least significant bits are to be used in the  
 multiplication.

20 In this embodiment, no additional time is required in order to  
 carry out the multiply subtract function. This embodiment  
 includes the selection between the most significant and the least  
 significant part of the Y data, and it requires 2 to 1  
 25 multiplexers as shown in Figure 4. This embodiment also includes  
 the multiply subtract feature, which only requires that the 2 to  
 1 multiplexers be changed to 1-hot 4 to 1 multiplexers. A 1-hot 4  
 to 1 multiplexer has 1 control signal for each input data. The 1-  
 hot 4 to 1 multiplexers have the same timing as the 2 to 1  
 30 multiplexers and therefore do not increase the timing.

Reference is now made to table 3 which shows how the CX-C2X-SGN  
 coding can be modified so as to be used in subtract and  
 accumulate functions.

[illegible]

10







5 in this embodiment. Thus  $CX'$  and  $SGN'$  are both generated by an XOR function between two primary inputs. They are generated at the same time. Accordingly, the multiply subtract function does not require any additional time.

10 In both of the above described modifications, the carry i.e. +1 of the expression  $-Y = \bar{Y} + 1$  has to be taken into account. In both embodiments, this is taken into account in the Booth coding of the first partial product  $P_0$ . This partial product is centred on bit  $Y_0$ . The three bits are therefore  $Y_1$ ,  $Y_0$  and  $Y_{-1}$ .  $Y_{-1}$  is  
15 assumed to have the value "0". As this bit is "0" only half of the possible values of the triplet can occur.

Table 4 below shows the possible values where  $N=0$ .

20

$Y_1$	$Y_0$	Partial product for $Y_1:Y_0$	$\bar{Y}_1$	$\bar{Y}_0$	$\bar{Y}_{-1}$ (='0')	Partial product for $\bar{Y}_1:\bar{Y}_0$	Partial product taking the carry into account
0	0	0	1	1	0	-X	0
0	1	X	1	0	0	-2X	-X
1	0	-2X	0	1	0	X	2X
1	1	-X	0	0	0	0	X

When calculating  $-P$ ,  $\bar{Y}_1, \bar{Y}_0$  and the carry need to be taken into account. The table above shows the partial product for  $\bar{Y}_1$  and  $\bar{Y}_0$  with  $\bar{Y}_{-1}$  being 0. The final column shows partial product that  
25 has to be generated when taking into account bits  $Y_1$  and  $Y_0$  and also the carry. Taking the carry into account (value: +1) means

5 adding  $1 \cdot X$  to the partial products of the previous column.

It can be seen from the above table that when calculating  $-P$ , the first partial product  $P_0$  is opposite of the partial product which is generated when calculating  $P_0$  for a multiply accumulate function.

Accordingly, the carry can be taken into account using the same mechanism as described hereinbefore for the other partial products. In other words NZN and NZP are exchanged for SMGL-NZP-NZN coding and SGN is inverted for CX-C2X-SGN coding.

Reference is made to Figures 7 and 8 which shows an alternative embodiment of the present invention. In this embodiment the input to the Booth coder is modified instead of the output. Firstly,  $Y$  is modified if it is necessary. When  $-P$  is required,  $Y$  is modified to  $-Y = \bar{Y} + 1$ . This modified value is then input to the Booth coders. The carry (+1) is also taken into account when  $-Y$  is used instead of  $Y$ . Input  $Y_{2i-1}$  of the Booth coder which receives  $Y_1$  and  $Y_0$  receives the carry. In this embodiment there are no cells between the Booth coder and decoder. Rather XOR gates are provided on each of the inputs to the Booth coder. The XOR gates provide an XOR function between  $Y_i$  and  $MUL\_SUB$ . The XOR gates generate  $\bar{Y}_i$  when  $MUL\_SUB$  is "1". When  $MUL\_SUB$  is "0", the XOR gates generate  $\bar{Y}_i$ . By modifying the inputs to the Booth coder, the correct output is obtained for the input values for both the multiply accumulate and the multiply subtract functions.

Figure 8 shows the booth coder which deals with the carry. When  $MUL\_SUB$  is 1, then  $-Y$  is used, and the carry (+1) has to be taken into account. When  $MUL\_SUB$  is 0,  $Y$  is used, and no carry has to be used (carry=0). In fact,  $MUL\_SUB$  equals the carry. Then it is this signal which is input on input  $Y_{2i-1}$  of the Booth coder receiving  $Y_1$  and  $Y_0$ .  $Y_1$  is connected to the  $2N+1$  input and  $Y_0$  is connected to the  $2N$  input. In an accumulate function input  $Y_{2i-1}$  is fed to 'C'. XOR gates are provided on the other two inputs which perform an XOR function between  $MUL\_SUB$  and  $Y_0$ , and  $MUL\_SUB$  and  $Y_1$  respectively.

In the embodiments described hereinbefore, XOR gates have been used. These gates can be replaced by an exclusive NOR gate or any other suitable logic. The values of some signals may need to be inverted.

Embodiments of the present invention are particularly applicable to digital signal processors and can be incorporated in integrated circuits.